NAVAL
POSTGRADUATE
SCHOOL

MONTEREY, CALIFORNIA

THESIS

PARTS-BASED DETECTION OF AK-47S FOR FORENSIC VIDEO ANALYSIS

by

Justin Jones

September 2010

Thesis Advisor:	Mathias Kölsch
Second Reader:	Simson L. Garfinkel

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE		3. DATES COVERED *(From — To)*
9–9–2010	Master's Thesis		2008-07-01—2010-09-28

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Parts-Based Detection of AK-47s for Forensic Video Analysis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Justin Jones	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Postgraduate School Monterey, CA 93943	

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
Department of the Navy	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. **IRB Protocol Number: n/a**

14. ABSTRACT

Law enforcement, military personnel, and forensic analysts are increasingly reliant on imaging systems to perform in a hostile environment and require a robust method to efficiently locate objects of interest in videos and still images. Current approaches require a full-time operator to monitor a surveillance video or to sift a hard drive for suspicious content. In this thesis, we demonstrate the effectiveness of automated analysis tools to detect AK-47s in images. By training on a large corpus of labeled data, we created Viola-Jones classifiers for detection of whole AK-47s and parts of an AK-47. Parts-based detections were then compared against learned models using support vector machines and multi-layer perceptrons. The results of this research show that parts-based classifiers combined with the above techniques leverage the high recall capability of part detectors and significantly reduce false positives in comparison to both the part and whole object classifiers. Techniques utilized in this thesis facilitate the creation of an automated capability for detecting AK-47s in support of the law enforcement and intelligence communities.

15. SUBJECT TERMS

Machine Learning,Parts Based Object Detection,Computer Forensics, Supervised Learning, VIDEX

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
					19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	UU	71	

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

PARTS-BASED DETECTION OF AK-47S FOR FORENSIC VIDEO ANALYSIS

Justin Jones
Captain, United States Marine Corps
B.S., Excelsior College, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 2010

Author: Justin Jones

Approved by: Mathias Kölsch
                           Thesis Advisor

                           Simson L. Garfinkel
                           Second Reader

                           Peter J. Denning
                           Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Law enforcement, military personnel, and forensic analysts are increasingly reliant on imaging systems to perform in a hostile environment and require a robust method to efficiently locate objects of interest in videos and still images. Current approaches require a full-time operator to monitor a surveillance video or to sift a hard drive for suspicious content. In this thesis, we demonstrate the effectiveness of automated analysis tools to detect AK-47s in images. By training on a large corpus of labeled data, we created Viola-Jones classifiers for detection of whole AK-47s and parts of an AK-47. Parts-based detections were then compared against learned models using support vector machines and multi-layer perceptrons. The results of this research show that parts-based classifiers combined with the above techniques leverage the high recall capability of part detectors and significantly reduce false positives in comparison to both the part and whole object classifiers. Techniques utilized in this thesis facilitate the creation of an automated capability for detecting AK-47s in support of the law enforcement and intelligence communities.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgements

First and foremost, I would like to thank Mathias Kölsch, whose encouragement and patience showed me the way and ensured that I completed this thesis. Thank you for the constant support.

To Dr. Garfinkel, thank you for the guidance and help with providing an application for this work.

I would also like to thank my good friend LT. Robert Zaborowski for his help with the prototype and in getting experiments to run on Hamming.

To my loving and supportive wife, Erin, who has always been dedicated to our family and my career. I could not have done this without you.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

Intelligence personnel supporting modern military operations are increasingly reliant on imaging systems to perform in hostile environments. From computer forensics to surveillance, imaging systems impact commander decision making. Correspondingly, increasing amounts of data must be processed to produce an analyzed intelligence product. With the explosion of UAV technology, wide area surveillance assets, and deployment of ground based surveillance platforms, intelligence personnel do not have the manpower to monitor all video feeds for real time decision making. Intelligence analysts also require mechanisms to detect suspicious material during collection operations in order to identify suspicious websites or other targets for monitoring. At the same time, ground forces must increasingly cope with exploiting useful data from electronic devices and media captured during raids on insurgent or terrorist safe houses.

This thesis demonstrates computer vision techniques for surveillance, computer forensics, and collection operations, by detecting suspicious objects in images. It focuses on the detection of AK-47s, due to their prevalence and potential for a variety of intelligence applications. Using computers to analyze images and video streams can decrease the response time to suspicious activity, allowing for real time alerts to be sent to forces and directly leading to lives saved and the disruption of enemy activities.

1.1 Operational Need

In recent years, new threats have emerged against the United States of America. These new threats can easily hide among the populace, thwart traditional combat intelligence gathering methods, and exploit seams in the authorities and capabilities of military and government intelligence organizations. With terrorist groups, insurgencies, piracy, drug cartels and other organized crime groups, as well as the emergence of potential peer competitors all posing significant threats to U.S. national security, intelligence professionals require new methods that flexibly support a variety of environments and facilitate rapid decision making.

In order to counter threats in these complex environments, forces must be able to precisely locate an individual operating within a large population. Compounding this problem is the amount of information entering an intelligence cell. In 2009 alone, Unmanned Aerial Vehicles (UAVs) from the United States produced 24 years worth of video, if watched continuously [1], with

new UAV models projected to increase data volume many times over. While UAV technology does provide the warfighter with significant advantages, more data does not necessarily equate to better information. The same applies to the volume of forensic materials collected by ground forces, and intelligence and propaganda intercepted by intelligence collectors. In order to be relevant to operational forces, the data must be processed, which is a significant weak point with modern intelligence mechanisms.

1.2 Computer Vision Support to Modern Military Operations

1.2.1 Intelligence, Surveillance, and Reconnaissance

Vision techniques offer substantial benefits for current and future Intelligence, Surveillance, Reconnaissance (ISR) systems. Full time operators are typically employed to monitor live feeds for direct support to operations. Using full time operators is inefficient, does not scale well, and may prove to eventually be infeasible with the increase in data feeds from new UAV models. In order to progress to an expansive persistent ISR system with the capability to provide real time warnings to front line troops, vision techniques identifying weapons can be incorporated with surveillance feeds. Reliable vision techniques can provide an operator with the capability to monitor more than one system and facilitate rapid decision making, potentially decreasing response time to an event.

1.2.2 Collection Operations

Image processing in support of collection operations can support insurgent/terrorist network targeting through identification of network composition, intent, potential targets, and associated mechanisms with the aim of disrupting an insurgent/terrorist planning cycle. Weapons can be found in a variety of insurgent/terrorist media (Figure 1.1) and thus can be used to focus intelligence collection efforts in order to find the proverbial "needle in a haystack."

1.2.3 Forensics

Detained persons must be evaluated and released if there is a lack of evidence implicating the individual. Due to the volume of detainees in a modern combat area, intelligence personnel must rapidly "triage" detainees and focus on those persons likely having knowledge of enemy activities. Raids on suspected terrorist safe houses typically produce a wide variety of electronic devices and media, which must be processed in a short period of time to support the interrogation process. Image processing techniques can be used to analyze the large volume of photos and videos that may be found in captured media, prioritizing those files for viewing by

Figure 1.1: Terrorist Media Often Contains Weapons That Can Be Used To Focus Collection Efforts. Image is Publicly Available at [2].

an intelligence analyst. Instead of having to view all images on a hard drive, an analyst can first be directed to those photos having suspicious items. By focusing the analytical process, valuable information can be rapidly provided to interrogators for use in determining a detainee's affiliation with enemy organizations, potential position, and likely activities.

1.3 Parts-based Object Detection Using Viola-Jones Classifiers

While object detection has always been a major focus of computer vision research, recent advances in the field have given rise to successes in a variety of real time applications. Of note, Viola-Jones classifiers have demonstrated particular successes in applications requiring face detection at a range of scales and have the capability to locate objects in video and still images. This rapid detection capability thus facilitates the development of a classifier that can be used in an array of intelligence applications.

While the Viola-Jones classifier has many desirable properties, detection of an object in a photo or video may be hindered by partial occlusion or backgrounds that reduce the silhouette of an object. A classifier trained on the whole object is less likely to respond as a positive if part of the object is missing. Parts-based techniques have the capability of increasing the rate of detection, as the simpler shape of a part will be more likely to positively respond. By first

detecting the parts, these detections can then be compared against a learned model to see if the part detections are consistent with a known geometry. Additionally, since each subwindow of an image is evaluated independently with a Viola-Jones classifier, false detections are common. By using parts-based techniques, subwindows can be evaluated individually by each part classifier, with the final object classification delayed until all part classifiers have been applied to the image, potentially leading to a decrease in false detections.

1.4   Research Questions

This thesis addresses the following research questions:

(a) Can a parts-based Viola-Jones classifier be effective in finding AK-47s in photos and videos?

(b) Can a parts-based Viola-Jones classifier have increased detection rates in comparison to a Viola-Jones classifier designed to detect the entire object?

(c) Can a parts-based Viola-Jones classifier have decreased false positive rates in comparison to a Viola-Jones classifier designed to detect the entire object?

In order to answer these questions, Viola-Jones classifiers were trained to detect the whole AK-47, as well as individual parts of the weapon. For the parts-based classifiers, a support vector machine and multilayer perceptron were used to develop a geometric model of the part configurations for comparison against each other, as well as in comparison against whole trained classifiers and part-only classifiers.

1.5   Results

The results of this research show that parts-based Viola-Jones classifiers combined with either a support vector machine or multilayer perceptron leverage the high recall capability of part detectors and significantly reduce false positives in comparison to both the individual parts by themselves and whole object detectors, when used with discriminative part classifiers. Classifiers trained to detect parts of an AK-47 exhibit a high recall, but a poor false positive rate when compared against classifiers trained on the whole object. Viola-Jones classifiers can be used to effectively detect weapons in video under a variety of lighting conditions and scales. Additionally, in-plane rotated AK-47s can be detected at a variety of angles by training object classifiers at a specific orientation, and then applying the classifier to rotated images.

4

1.6 Organization of Thesis

The thesis is organized as follows:

(a) Chapter 1 discusses the modern operational environment and need for computer vision techniques in support of intelligence activities.

(b) Chapter 2 contains related work relevant to parts-based object recognition.

(c) Chapter 3 discusses the methods selected for military applications, techniques for the development of Viola-Jones classifiers, Support Vector Machines, and Multilayer Perceptrons, and procedures for the creation of a parts-based structural model for the detection of AK-47s.

(d) Chapter 4 contains experiment design and sources of data.

(e) Chapter 5 contains results and analysis of experiments.

(f) Chapter 6 contains concluding remarks and possible future areas for research that exploit results from this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Prior and Related Work

Object detection is important in many applications in the business, science, and intelligence fields. Over time, numerous methods have been developed to identify objects in both still images and video. Some methods rely on color and may be somewhat scale and rotation invariant, while others may be based on pixel, shape, or edge detections from a variety of filters. Still others may use motion to classify the object or a set of actions. The method is typically chosen in the light of the given task. Color may not be an option given the purpose of the proposed system, while others that are effective may not be fast enough to implement in a real time detection system. This section compares contemporary computer vision techniques for identifying objects and sets the context for the work in this thesis.

2.1 Related Work

A number of feature types, feature selection mechanisms, and parts combination techniques have contributed to the goal of identifying objects in still images and video frames. This section will compare contemporary techniques with the methods chosen in this paper.

2.1.1 Origins of Viola-Jones Classifiers

In 2001, Paul Viola and Michael Jones proposed a method of detecting faces based on Haar wavelets, trained with Adaboost, and combined in a sequence they called a "cascade of features [3, 4]. Using only upright rectangles that scaled both horizontally and vertically in constant time, they achieved detection rates of 77.8% while at the same time achieving only 5 false positives in a test of 23 images with 149 faces [4]. The authors also demonstrated that their classifier was capable of detecting non-rotated faces at a speed of 15 frames per second [3], offering the potential for real time object detection for video applications. Rainer Lienhart and Jochen Maydt then expanded the work of Paul Viola and Michael Jones by developing a richer feature set that included 45-degree rotated features for use in training a strong classifier [5]. Both of these techniques rely on supervised learning to annotate the object in an image for training. Since the Viola-Jones classifier is not rotationally invariant, techniques for developing rotated Haar features with the potential to locate in-plane rotated objects in an efficient manner have been created [6]. The Viola-Jones method is appropriate for training whole and part de-

tectors, and was chosen in this paper for it's speed of detection, as well as the low false positive rate.

The standard Viola-Jones cascade is binary and is executed independently on each subwindow of an image. The number of subwindows in an image can be very large, and may result in false positives being generated at an unacceptable rate. Additionally, it is possible that a subwindow of an image containing an object may not make it completely through a cascade due to the environment. In order to improve the binary cascade, a "fuzzy" framework can be provided through the use of a voting procedure over local subwindows for cascades that do not make it completely to the end [7]. Methods for a boosted classifier with the ability to convert raw classifier outputs into posterior probabilities also exist and are capable of evaluating an object's likelihood distribution over a local area of an image [8].

2.1.2 Histograms of Oriented Gradients(HoG)

Other parts-based approaches rely on a variety of features. Histograms of Oriented Gradients (HoG) are a common approach where orientations of gradients are summed in a portion of the image [9, 10, 11, 12, 13, 14]. Since histograms are calculated over local regions, the method is somewhat invariant to geometric and photometric lighting changes [10].

2.1.3 Eigenfaces

Eigenfaces, or the use of principal components analysis to find the vectors of pixel features with the largest eigenvalues of a face, have also been used in a variety of approaches [15, 16, 17]. This method is global, works well for face recognition and when lighting variation is small, but performance deteriorates as the lighting variation increases [15]. Independent Component Analysis (ICA) is a technique that that can better compensate by separating a multivariate signal into subcomponents, and thereby determine independent directions in the feature space, rather than the dominant ones detected by PCA. This technique is better suited for classification tasks than PCA.

2.1.4 Edge Detection

Edge features are found in variety of methods in parts-based object detection. Gabor filters are used for edge detection in a number of related work [18, 19, 20] and are well suited for representation of textures [21]. Gabor filters are Gaussian kernels that have been modulated by an oscillating plane wave [21], and whose input response (the original pixel value) is determined by its location and value with respect to the Gaussian kernel, modulated by the waves parameters

(orientation, wavelength, phase offset, etc.). This creates distinctive activations for objects at particular spatial locations, and can also be used to create a sparse object representation [21].

Shape recognition can also be conducted through the use of a Canny edge detector [22]. Canny edge detection can be noise sensitive, and is typically conducted after convolving an image with a Gaussian filter. Based on the first derivative of a Gaussian, four filters that detect vertical, horizontal, and diagonal edges are applied to the Gaussian blurred image. Values from the first derivative of the responses in the horizontal and vertical directions can then be used to determine the edge gradient and direction. Canny edge detection has been shown to have a bias towards both horizontal and vertical edges and does not provide a good approximation of rotational symmetry [23].

Edge detection based on gradient approaches through the use of the Laplacian operator have also been applied in related works [24, 22]. The Laplace operator is useful for blob detection, and is found by the sum of differences over the nearest neighbors of the central pixel, after convolving with a Gaussian kernel. The operator responds with a strong positive reaction to dark blobs, and a strong negative reaction to bright blobs [25]. One of the disadvantages of this approach is the operator response is dependent on the size of the Gaussian kernel for pre-smoothing and the size of the blob structures. A multi-scale approach is thus required to find blobs of an unknown size [25].

2.1.5 Interest Points

Interest point detection is a common approach to finding and tracking objects in photos and video. Typically, interest point detection is used to find "corners", or areas of an image that have gradient changes in multiple directions. Interest points are somewhat stable to affine transformations, scale changes, and rotations/translations and are suitable for localizing an object in an image [26]. Interest point detectors are a commonly used in object detection [27, 28, 14, 29] and can be used for unsupervised part learning.

2.1.6 Part Learning Strategies

After choosing a feature set, the best features must be selected out of all the features for learning. Strategies for part learning vary in accordance with the overall goals for object detection. Some strategies emphasize that parts are clustered into feature sets that are as different as possible, which provides for learning the abstract idea of a part [27]. This is accomplished by a cost function which evaluates a part based off of normalized correlation and attempts to place similar

parts into the same bin with a feature id. Part learning techniques using clustering can also be used to place parts into a tree for fast object retrieval [30].

Other methods seek to select those features that best classify a validation set. Statistical boosting is a common mechanism used to find and train those features that best classify positive and negative examples [3, 4, 24, 19, 20, 9]. The Viola-Jones method uses a form a statistical boosting called Adaboost to simultaneously select and train the weak classifiers composed of Haar features.

Rather than choosing a part based on its abstraction, or ability to best classify a validation set, other techniques use part learning for the final classification performance of an object. In [28], local part information is used in conjunction with global cues about an object's silhouette. Object detection based off part appearance and then refined through geometric location can lead to inaccuracies if the part appearance is noisy or ambiguous [14]. This technique seeks to label parts by appearance and location simultaneously through the use of a random field framework. In [13], a technique for maximizing over latent part locations is used to determine the presence of a whole object, as opposed to a discriminative process where an object is determined to not be there if enough criteria is not met. Random local feature sampling can also used for part learning. With this method, random part sampling is matched to randomly trained parts [18] through a response to Gabor filters and a Euclidean distance measurement of the local maxima responses.

2.1.7 Modeling Part Combinations

After choosing a feature set, and selecting the best set of features to learn, the location of parts should be combined based on a trained model. By using the geometry of part detections, the number of false detections and the amount of feature space to search can be reduced [24]. The Sparse Network of Winnows (SNOW) architecture can be used to learn associated distance and direction combinations between parts at various scales [27]. This technique can be used for parts-based learning, as in any image, it is likely that only a small subset of the features are present. This technique is suitable for use in applications with spare feature representations. Gaussians can also be used to model the location of parts with the detection of one part improving the likelihood and location of detecting another part and can be used to detect both rigid and flexible objects [29]. Gaussians are used in a variety of related works [24, 13, 29, 31] to model part combinations.

Markov random fields are also used for modeling spatial part locations as an undirected graph representing the dependencies between detected parts [18, 14]. Markov random fields can represent cyclic dependencies, where a Bayesian network cannot [32], and are useful for determining the joint probability of parts co-located in a graph.

In a related work, object recognition is conducted through a mixture of multi-scale part models [13]. Parts combinations are learned through the use of a star based topology that applies a root filter at a lower resolution, then parts-based filters at twice the spatial resolution, with the total score for detection being a combination of root filter detection location, parts detected, and their associated locations in relation to the trained model. Latent SVM is then used to train the parts models on partially labeled data. Support Vector Machines are also used in [19, 20] for use in discriminating the location of part detections.

In another related work involving parts-based pedestrian detection [28], the spatial relationship among detected parts was represented by having extracted patch sections compared to a codebook via normalized grey scale correlation and then having each matched codebook entry "vote" for the probable location of the object. The probablistic vote is based on the probability of matching a codebook entry, with each code book entry having a corresponding probability distribution for the center of the object. The object center was then found in the 3D voting space by searching for maxima with Mean Shift Mode estimation.

Facial detection with local feature sampling can also be conducted by a neural network based approach [33]. Three types of "hidden" units, four which evaluate 10x10 pixel subregions, sixteen which evaluate 5x5 pixel subregions, and six which evaluate are 20x5 pixel horizontal stripes, are used to detect local features. A neural network based filter takes the responses from the hidden unit pixel values and values from the horizontal strip regions, and outputs a result for the window being scanned. Multiple detections spanning across windows are then combined into a single detection.

2.1.8   Automated Part Training

Automated part training has some distinct advantages over just hard coding a part to learn. First, automated part learning decreases the amount of time required to label data, as a person may only be required to label the object, instead of having to laboriously label every part, or trace an outline around an object. Additionally, the best set of parts to learn might not be immediately clear, so a trial and error basis may produce better results than having a human choose the parts

11

to learn. Unsupervised part training is conducted in a number of related works [13, 18, 14, 27]. Unsupervised part clustering is also used in a variety of related works to collect related features into a single group [14, 27, 28, 30]. By first identifying an area (typically through the use of an interest point detector), patches around the interest point can be extracted, and similar patches can be clustered into a "vocabulary" [27] that composes the object. While automated training has numerous advantages, manual part labeling is still a common approach [33, 24].

2.1.9   Suspicious Behavior Recognition

A number of other techniques have also been used to identify potentially suspicious events or for human surveillance in video applications. Automated video analysis techniques for finding violent events or surveillance of humans in video have been explored in detail in a variety of related works [34, 35, 36, 37, 38, 39]. The methods in these seminal papers rely on audio visual cues, such as the sudden flash or sound of an explosion, dynamic changes between frames, or motion accelerations in relation to human silhouettes to determine the presence of a violent event. While suitable for video, these techniques are not likely to be as applicable to forensic applications searching still frames for suspicious objects.

CHAPTER 3:
Methods For Detecting AK-47s In Support of Military Applications

This chapter details the selection of methods in this thesis and applicability to modern military operations. Since Viola-Jones classifiers are a foundational technique, Haar features, the integral image, and the binary cascade are discussed in some detail to provide background to the reader. This chapter also discusses the general development of Support Vector Machines and Multilayer Perceptrons and their utility for final classification of part combinations. Finally, the creation of a novel parts-based structural model for AK-47s is also discussed in detail.

3.1 Selection of Techniques for Weapon Detection in Support of Forensics and Surveillance Applications

Given the amount of related work in object detection, there are a variety of techniques to choose from for identifying weapons in an image for surveillance or forensics applications. While color based learning of an object has significant advantages due to its scale and rotation invariance, it is likely not well suited for use in support of military operations where object identification must be conducted in low light conditions (including images via night vision devices). Additionally, given that the classifier must also be able to locate objects in still frames as well as video, techniques that rely on motion, sound, or frame differencing for classification are also not likely to be appropriate. Techniques that are appropriate for real time surveillance monitoring would be also be suitable for scanning images or subsampling video for suspicious objects in a forensics application.

Since speed, detection at a variety of scales and lighting conditions, and support for still images and video are primary requirements, Viola-Jones classifiers offer some of the best potential for employment in military applications. Though weapons are rigid objects, which are well suited for a template based approach, one issue does arise when trying to locate them in images. Weapons typically are recognized by silhouette, and therefore most do not have much internal structure that makes maximum use of the Haar feature set. Additionally, background objects and occlusion can significantly change an objects shape [22], thereby making a classifier trained on the entire object less likely to respond as a positive detection.

Parts-based learning offers some benefit for detecting objects under "noisy" conditions over trying to detect the whole object. Parts of an object have fewer information than the whole object. This means that a classifier trained on a part is also more likely to respond to this simpler object, since there are likely less conditions that must be met in order for the classifier to respond as a positive detection. On the other hand, since the detector is more likely to respond, there is likely to be a significant increase in the amount of false positives generated by a part trained classifier. By using multiple part detections, as well as their relative spatial geometry, recall may be increased, while at the same time keeping false detections to an acceptable level.

3.2   General Machine Learning Techniques Utilized in this Thesis

This section discusses the general machine learning techniques required to develop Viola-Jones classifiers, Support Vector Machines, and Multilayer Perceptrons. Viola-Jones classifiers are used as the base technique to recognize whole AK-47s, as well as the individual components for employment in the parts-based methods. In the parts-based approach, Support Vector Machines and Multilayer Perceptrons provide a final classification for detected part combinations.

3.2.1   Viola-Jones Classifiers

Viola-Jones classifiers are a well known technique for locating objects in videos and photos. Fast and efficient, Viola-Jones classifiers are easily trained on a corpus of positive and negative image samples. This is a supervised learning technique, where a human must first annotate images with a bounding box for the object. The positive image is then cropped to the bounding box location, converted to grey scale to eliminate the influence of color, and normalized to a user specified size. While better classifiers do exist, the primary advantage of a Viola-Jones classifier is the speed and efficiency with which it can detect an object. Viola-Jones classifiers can be run in real time on video streams to detect objects of interest. The following subsections explains the underlying functionality of a Viola-Jones classifier.

Haar Features

The Viola-Jones classifier uses features that are based on the concept of Haar wavelets, which is a square-integral function for approximating continous functions. This square wave has the properties of a regular wave, in that there is a repeatable high and low amplitude, with a fixed wavelength. For image detection, this is exactly represented by a two-dimensional pair of adjacent rectangles. One of the adjacent rectangles has a light area, and the other is dark, which

represents the rise and fall of the Haar wavelet. When the pair of rectangles is placed over an image, the average value of the dark region is subtracted from the average value of the light region [40]. This weak classifier indicates that a feature is present if the resulting subtracted value is greater than a threshold found by training over a corpus of positive and negative images. The advantage of this feature set is the extreme efficiency with which it can detect a feature. In an example involving face detection with Haar features (Figure 3.1), the first set of rectangles respond to the fact that the eye region is darker than the cheekbones, while the second set of rectangles responds to the nose being lighter than the eyes. The original Viola-Jones classifiers incorporated a set of non-rotated rectangles, but was expanded with 45-degree rotated rectangles to provide a richer feature set for learning. In the total set, there are 8 line features, 4 edge features, and 2 center surround features, which can be combined in a linear combination as a strong classifier (Figure 3.2).
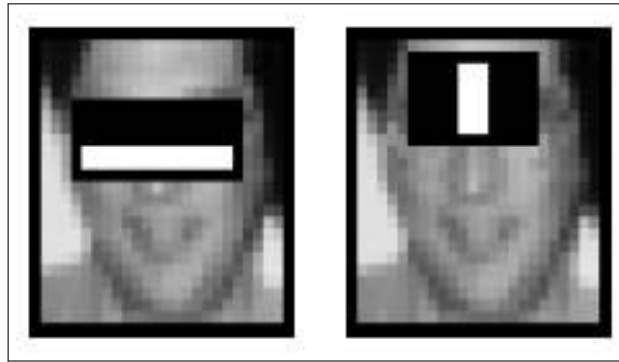

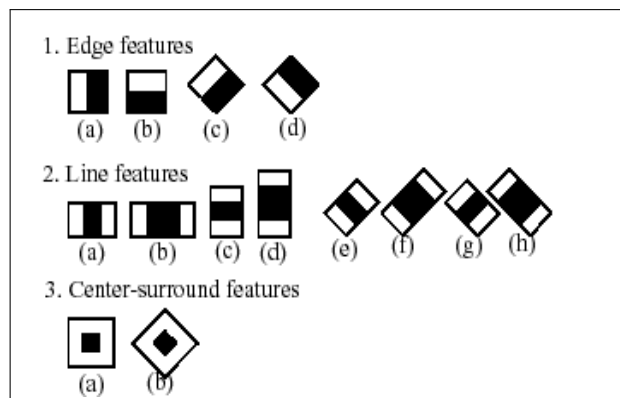Figure 3.1: Haar Features for Face Detection. Image from ([3]).

Figure 3.2: Haar Features. Image from ([5]).

15

The Integral Image

Since the weak classifier is determined by the difference of the sum of the values in the light region and the sum of the values in the dark region being over some threshold, a quick mechanism is needed to determine the values in these regions. Just as the concept of integrating continuous functions entails the summing of rectangles to determine the area below a curve, the integral image is made by summing all the pixels above and to the left of some x,y pixel value, with the value at the location x,y being inclusive (Figure 3.3) [40, 4, 3]. This is an extremely efficient technique for finding the average pixel value in an area of an image. All that is required to find the pixel value for any upright rectangle in any area of an image is four table lookups then dividing by the area of the rectangle. When determining the value of a 45 degree rotated rectangle, two passes are required over all the pixel values, once from left to right and top to bottom, then from right to left and bottom to top [5]. The rotated rectangle value at a pixel location x,y can then be calculated through 4 table lookups. When using a Haar feature, this is done for both the light and the dark rectangles, and the difference of summed values is compared against a threshold to determine if the feature is present.
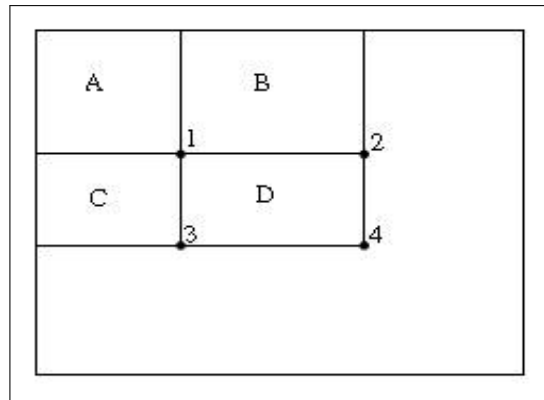
Figure 3.3: Integral Image. Image from ([3]).

(a) The value of pixel at x,y is the sum of all pixels above and to the left, with x,y inclusive. The value of any rectangle in the image can be computed with a total of 4 table lookups. To get the value in rectangle D, take the large rectangle (A+B+C+D), and subtract out the areas that are not used to compute the integral (namely rectangles A+C and A+B). The sum of pixel values in D then is pixel values at 4+1-(2+3).

Adaboost for Learning a Strong Classifier

The Haar feature by itself constitutes a weak classifier, which is a classifier that gets the right answer just slightly better than random chance [40]. When a number of weak classifiers are

combined in a series, then an overall effect is produced that is much stronger than any one feature by itself. Paul Viola and Michael Jones chose the Adaboost algorithm for its ability to select and simultaneously train a set of weak classifiers [4, 3]. Using the Haar features as weak classifiers, the Adaboost algorithm iteratively runs each feature over a set of positively and negatively labeled images. The feature (pair of light/dark rectangles) that best classifies the two examples correctly (i.e. has the lowest error) is then selected to update the weights. When updating the weights, incorrectly classified examples are given more weight than those that are correctly classified. This is an important point, in that when learning, it is usually the marginal examples that provide the best examples for learning a new concept. Examples that are "black and white" can be classified quite easily. It is the "grey" areas that probably provide the best examples. After choosing the best feature and updating the appropriate threshold, the distribution of weights is recomputed, and the cycle continues until an appropriate threshold has been reached.

Cascade of Classifiers

Due to the large amount of subwindows checked in an image, a fast and efficient method is required to achieve speeds that enable real time object detection. Using the Adaboost trained classifiers as filters, the filters are combined into a degenerate tree, where the branch is a binary classifier indicating a positive or a negative detection [40, 3]. The classifiers are arranged in such a way that the simpler classifiers that detect most positive instances, but reject many of the subwindows are called first, then more complex classifiers are called in order to keep false positives lower [4, 3]. This method ensures that the majority of subwindows in the image that do not contain the object of interest are quickly evaluated and passed (Figure 3.4).
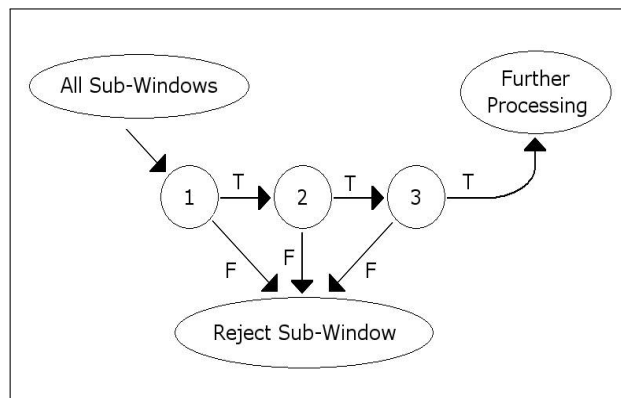

Figure 3.4: The Viola-Jones Cascade of Adaboost Trained Filters. Image is publicly available at ([41]).

17

3.2.2 Combining Classifiers with Machine Learning Techniques

The above techniques are suitable for training either whole or parts-based classifiers to identify an object. Using various machine learning techniques, the locations of part detections can then be checked against a learned model. Support Vector Machines and Multilayer Perceptrons were used in this thesis to provide final classifications in relation to a structural model learned from a training image set. The general development of Support Vector Machines and Multilayer Perceptrons are outlined below.

Support Vector Machines

Support Vector Machines are a fast and efficient way to classify the parts detected by the Viola-Jones classifiers. Using a feature vector of just 3 dimensions, a support vector machine can be trained to validate the geometry of the part detections from the Viola-Jones classifiers.

Given a feature vector of object attributes as points in space, support vector machines attempt to find a hyperplane to separate the two classes [42]. A good hyperplane is one that maximizes the distance to the classified examples that are closest to the plane. See Figure 3.5. If the data is linearly separable, then this hyperplane is called the maximum-margin hyperplane. Just as noted above in the Adaboost chapter, the marginal examples for both classes have the greatest chance of being misclassified. This is precisely where the support vector machine creates a hyperplane, and these marginal examples are known as the support vectors. For an n-dimensional feature vector, SVM attempts to find an $n - 1$ dimensional plane to classify the two examples. If the data is not linearly separable, then a hyperplane may be found by a kernel function that projects the data into a higher dimensional space or through the use of "slack variables" that allow for a soft margin that accounts for misclassified examples [42].

Multilayer Perceptrons

The Multilayer Perceptron (MLP) is one of the most common types of neural networks. It is a feed forward neural network that maps a set of input data from a provided feature vector to an output vector [44, 42]. This is a supervised learning technique. The MLP consists of a minimum of 3 layers: the input layer, one or more hidden layers, followed by an output layer (See Figure 3.6). Taking in an input vector, weights on the links for the connections from each perceptron are learned via backpropagation. By iteratively working backwards from the desired outputs, weights can be determined for the links between the layers. Since the desired output is known, and the sampled output from the network is also known, it is possible to compute the local error for each output neuron [45]. The local error is a factor with which the output of the neuron
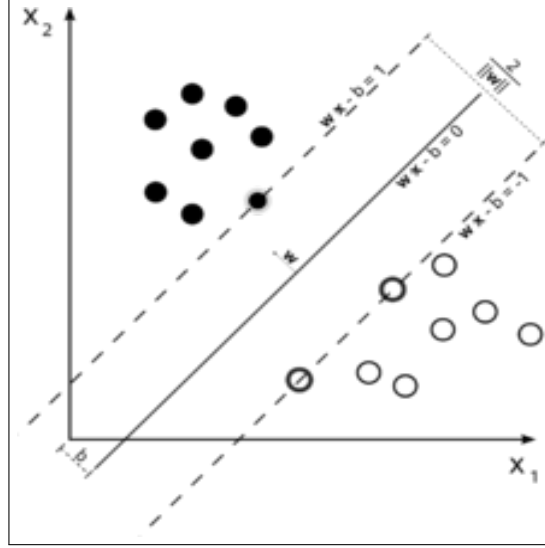
Figure 3.5: Maximum Margin Hyperplane. Image is Publicly Available at ([43]).

must have been to match the desired output. The weight of the neuron is then adjusted to the local error in order to compensate for the incorrect classification. Neurons on the previous level are then assigned "blame" for the local error caused at the current level. Neurons with stronger weights at the previous level receive more "blame" for the local error caused at the current level. The step is then repeated for the neurons at the previous level, with the "blame" as its local error [45]. After iteratively computing the error at each level, a set of weights are created for each link for each neuron between the layers. New unclassified feature vectors can then be provided to the trained network, with each non-linear activation function determining whether a neuron fires, with a corresponding weight for the link. After traveling through the entire network, an output is provided with a total classification of the input vector.

3.3 Structural Model for Parts-based AK-47 Detection Using Viola-Jones Classifiers

This thesis incorporates a novel technique for the creation of a structural model for AK-47s using parts-based Viola-Jones classifiers. This structural model assumes that parts detected belonging to an object are likely to be detected at a similar scales. The structural model also assumes that AK-47s are consistent with the training set: barrel pointing right, with in-plane and out-of plane rotations of no more than approximately 10 degrees. These assumptions allow for training a model in a specific configuration, which can then be used to find AK-47s in other
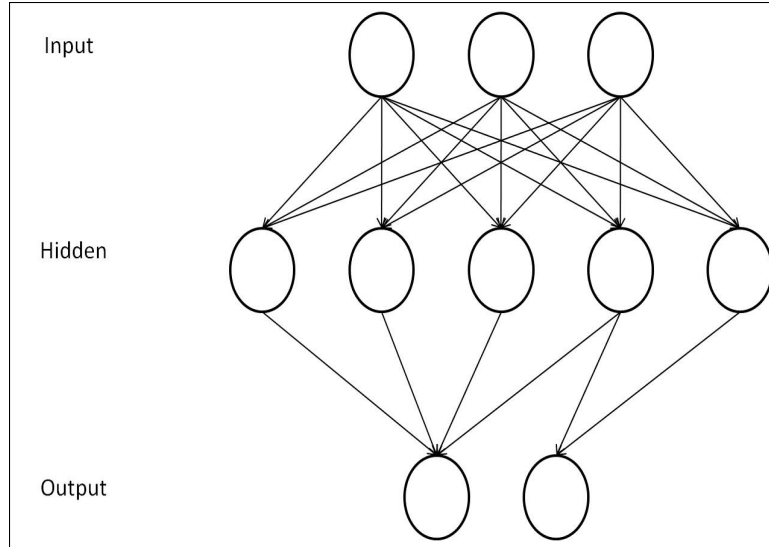
Figure 3.6: Multilayer Perceptron. Image is Publicly Available. References ([44]).

Figure 3.7: Sample Training Images for Left and Right Parts. Image is Publicly Available at [46].

orientations by rotating the image.

This structural model incorporates two parts. The rear end of the AK-47 encompassing the pistol grip and magazine is designated as the Left Half of the AK-47. The rifle stock was not used for training due to occlusion in a number of images, as well as due to the large number of varieties found in the training set. The Right Half of the AK-47 includes the hand guard, barrel, and sight post (Figure 3.7).

3.3.1   Radius of Object Detection

Since the Viola-Jones method evaluates subwindows of an image, a rectangular area containing the object is returned with the center of detection, as well as the width and height dimensions of the associated subwindow. In order to make a more efficient representation of the object

20

location for use in the following relative feature vector, the rectangular area is converted to a circle with the same center of detection, but with an associated radius (Equation 3.1).

$$DetectionRadius = (RectangleWidth + RectangleHeight) * 0.25 \qquad (3.1)$$

3.3.2   Structural Model Feature Vector

The structural model of the AK-47 provides a mechanism to ensure that parts detected by the left and right half AK-47 classifiers match a geometry consistent with the presence of an object. Parts detected are placed into a vector consisting of 3 elements:

(a) Difference between the left half detection center x value and the right half detection center x value normalized by the mean radius of the 2 detections (Equation 3.2).

$$NormalizedXDifference = \frac{(RightCenterXValue - LeftCenterXValue)}{((LeftRadius + RightRadius)/2)} \qquad (3.2)$$

(b) Difference between the left half detection center y value and the right half detection center y value normalized by the mean radius of the 2 detections (Equation 3.3).

$$NormalizedYDifference = \frac{(RightCenterYValue - LeftCenterYValue)}{((LeftRadius + RightRadius)/2)} \qquad (3.3)$$

(c) Difference between the left and right radii normalized by the mean radius of the two radii (Equation 3.4).

$$NormalizedRadiusDifference = \frac{(RightRadius - LeftRadius)}{((LeftRadius + RightRadius)/2)} \qquad (3.4)$$

By normalizing over the mean radius of detections, a feature vector is produced that accounts for the relative distance between the left and right half detection centers in both the x and y dimensions of an image (Figure 3.8). The normalized radius difference accounts for the assumption that left and right half detections should be at approximately the same scale.
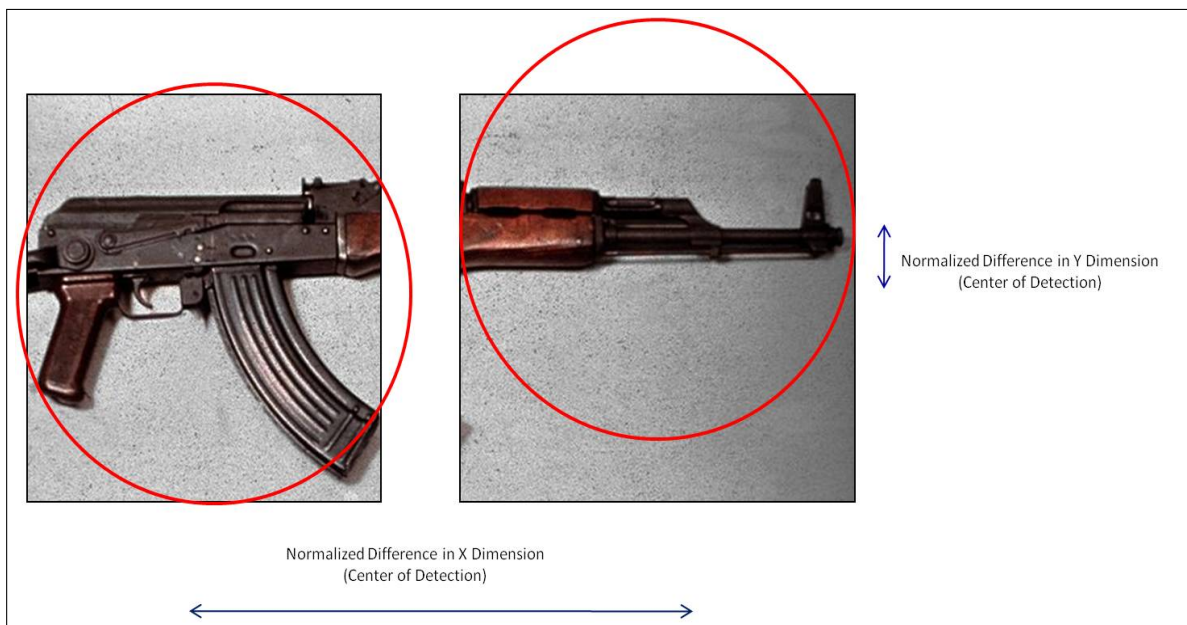
21

Figure 3.8: The AK-47 Structural Model. Image of AK-47 is Publicly Available at [46].

CHAPTER 4:
Experiments

This chapter describes the experiments conducted to train and test whole, part, and parts-based classifiers developed to detect AK-47s. Two Viola-Jones classifiers for the whole weapon were created to verify that a Viola-Jones classifier could be trained to find AK-47s in images and to establish a baseline for comparison against recognition through individual parts and parts-based techniques. Two separate part detectors were also trained, one for the rear end of the AK-47 and one for the barrel and front sight. These experiments were conducted to determine which part classifier is more discriminative in detecting AK-47s and to test the hypothesis that detection of an object through part recognition increases recall and the false positive rate in relation to a whole object. Finally, a novel approach developed in this thesis tests the hypothesis that a parts-based technique can combine the benefits of rapid Viola-Jones classifier detections with the increased recall capability of part detections, while simultaneously maintaining a lower false positive rate than classifiers trained on the whole AK-47. Classifier stages were also removed from individual classifier cascades to evaluate the impact of stage removal on recall and the false positive rate.

4.1   Sources of Training Data

All images for training and testing were provided by selecting frames from videos, which were obtained by searching the Internet. The strategy for training a classifier was to provide a number of images with AK-47s, all pointing right, with in-plane and out-of plane rotations of no more than approximately 10 degrees. The classifier was trained to recognize AK-47s in this specific orientation. Images were rotated and flipped to recognize AK-47s in other orientations. For the negative image set, images without AK-47s were used for training, including crowded city streets, villages, as well as people holding objects, so that a classifier would not inadvertently learn the hands of people.

4.2   Division of Data

Of the 18 total videos, 13 videos with 1146 cropped images of AK-47s were selected to train the classifiers. These images included a variety of backgrounds and configurations of the weapon. Weapon configurations included standard 30 and 40 round AK-47 magazines, with and without slings, and standard and pistol grip foregrips of various colored textures and materials. Due to

the wide variety of AK-47 stock types available and occlusion in a number of images, the rifle stock was not used for training. Images of AK-47s cropped for the training set included the pistol grip on the rear end of the AK-47 to the end of the barrel and front sight (Figure 3.7). For the negative training set, 5660 frames were split from 23 dynamic videos. A separate image set was developed for testing purposes. In this test set, 687 images containing AK-47 shooters were split from 5 videos. For the negative test image set, 7045 frames without AK-47s were split from 24 videos.

4.3   Normalization of Training Images

During the annotation process, a rectangular area containing the object was extracted from each positive image. Prior to training, annotated sections were normalized to a specific size and converted to grey scale. For classifiers trained on the whole object, annotated sections were normalized to 20x40 pixels. Parts were created by taking the annotated section and dividing the width in half. These images were then normalized to 20x20 pixels each before training (Figure 3.7).

4.4   Number of Images for Training

When training classifiers, more training data is typically better. The first whole classifier, designated Whole_AK, was trained with the 1146 positive and the default 2000 negative examples per stage. The next whole classifier, designated Whole_AK_Negative_Resistant was trained with more negative samples in order to improve the false positive rate. Whole_AK_Negative_-Resistant was trained with the 1146 positive samples and 5660 negative samples per stage. The Left_Half_Detector and Right_Half_Detector were trained with the same images as the Whole_-AK_Negative_Resistant classifier.

4.5   OpenCV Training for Viola-Jones Classifiers

After preparing the images, OpenCV's Haar training utility produced the classifiers specified above. A complete overview of the boosting process is contained in [3]. All classifiers were trained with the extended Haar feature set in non-symmetric mode, with each classifier's cascade containing 20 stages. The specified minimum hit rate for all classifiers was 0.998 per stage with a maximum false alarm rate for each stage of 0.5.

4.6 Training the SVM and MLP for Classifying Part Detections

A Support Vector Machine and a Multilayer Perceptron were independently trained to test consistancy of the detected part locations against the structural model and confirm the presence of an AK-47. In order to train the SVM and MLP to classify positive and negative instances of AK-47s, left and right half classifiers were applied to the 1146 positive training images, and 5560 negative training images, with post processing turned on. All combinations of left and right detections in a photo were kept. Detections inside the annotated box of the training set were considered to be true detections, while detections outside of the annotated box or in a negative photo were considered to be false detections. For a description of the structural model and vector used for training, see Chapter 3. A graph of the geometry of the detections is provided in Figure 4.1, showing the cluster of positive detections versus negative detections. Note that the normalized radii difference is not included in the graph. A hyperplane separating the positive and negative sets was then found with SVM. A Multilayer Perceptron was also trained on the same data set for comparison and in the case the data set was not linearly separable.
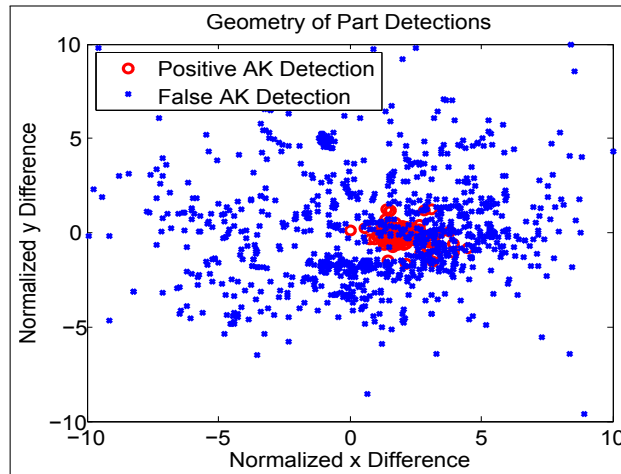

Figure 4.1: Plot of Normalized Relative Part Detections In the X and Y Dimensions After Applying Left and Right Classifiers to the Training Set.

(a) The red cluster represents relative part detections of actual AK-47s. These part detections were used to train a Support Vector Machine and a Multilayer Perceptron to classify detections against the AK-47 Structural Model.

4.7 Performance Measures

Performance was evaluated with two standard metrics. Temporal processing was not conducted to achieve the reported results. Temporal processing could increase the recall over a video sequence, since an AK-47 might not be detected in every frame, but possibly in subsequent frames. Recall, which is a traditional measure of completeness, measures the percentage of weapons detected in the image set (Equation 4.1). The false positive rate complements the recall measure for each classifier by providing the probability of a false detection for each subwindow evaluated (Equation 4.2). The following equation was used to determine recall for each classifier:

$$Recall = \frac{NumberOfWeaponsDetectedInSet}{TotalWeaponsInSet} \tag{4.1}$$

The following equation was used to determine the false positive rate:

$$FalsePositiveRate = \frac{NumberOfFalsePositivesInImageSet}{TotalAreasChecked} \tag{4.2}$$

While the test set contains images of many sizes, a secondary performance metric was devised to provide a more intuitive explanation of the false positive rate in relation to a simulated surveillance system. This performance metric is compared against a standard video with a frame size of 640 x 480 pixels, running 15 frames per second for one minute. This is a total of 900 frames per minute with several hundred thousand subwindows checked per frame. The total number of predicted false detections per minute is then the false positive rate times the amount of subwindows checked per minute of video (Equation 4.3).The following equation gives the number of predicted false positives per minute of video:

$$PredictedFalsePositivesPerMinute = FalsePositiveRate * AreasPerFrame * 15 * 60 \tag{4.3}$$

For a 640x480 image, a 20x40 whole trained detector is scanned over 314,319 areas. For the smaller 20x20 part detections, 352,718 areas are checked in each 640x480 image. For the entire test image set, 877,639,110 negative areas were checked for each whole detector, with 1,033,439,228 negative areas checked for each part detector.

4.8 Improving Recall by Reducing the Number of Stages

In order to improve the recall of a trained classifier, stages can be removed from the Viola-Jones cascade. While this makes the overall whole or part classifier more likely to indicate that an

object is present, there is also an increase in the number of false positives that will be detected by the classifier. Classifier stages were removed from the Viola-Jones cascades from each classifier until a recall above 85% was obtained. ROC curves for all classifiers are provided in Chapter 5. Tables of results are provided in Appendix A.

4.9   Detecting AK-47s in a Test Image with the Structural Model

After training Haar classifiers for the left and right halves of the AK-47 and training an SVM and MLP to classify part geometries in relation to the structural model, classifier combinations can be utilized to identify AK-47s in images. The left and right half detectors are first scanned over an image producing vectors containing the normalized x center difference, normalized y center difference, and normalized radii difference for all combinations of left and right detections. In general, Left_Half_Detector detections are much more discriminative than those generated by the Right_Half_Detector. Each vector is then evaluated with the support vector machine or a multi-layer perceptron, producing a final classification for the detection.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
Results and Analysis

This chapter presents the results from the experiments conducted for all classifiers and classifier combinations. First, whole trained classifiers provide the baseline for comparison against the other classifiers and methods employed in this thesis. Next, part classifiers show individual left and right half classifier performances and aid in determining which classifiers are more discriminative. Finally, parts-based techniques using the Viola-Jones part classifiers and either a Support Vector Machine and a Multilayer Perceptron are presented. All classifier perfomances are evaluated with Receiver Operating Characteristic (ROC) curves to evaluate the benefit (increase in recall) against the cost (higher false positive rate) caused by the removal of classifier stages from the Viola-Jones cascade. All graphs were produced in Matlab. Tables are included in Appendix A for additional information when referencing graphs. All source code and training and test corpora are available upon request.

5.1 AK-47 Detection with Whole Trained Classifiers

Two classifiers were separately trained to verify that Viola-Jones classifiers could detect AK-47s in images. The Whole_AK_Resistant classifier was trained with an increased number of images without AK-47s in order to test the hypothesis that training with more negative images could develop a classifier more resistant to false positives. Training for each classifier required approximately two days on a Intel Core 2 CPU at 2.4 GHz with 2 GB of RAM, with the Windows XP operating system.

5.1.1 Whole Trained Classifier Results

Results for the image set with the Whole_AK and Whole_AK_Resistant detectors are shown in Figure 5.1. For the whole trained detectors, Whole_AK (all stages) had a recall over the image set of 67.8%, but 67.6 false positives per minute of video (FPM). Whole_AK_Negative_Resistant (all stages) had a starting recall of 61.8%, but generated 41.2 FPM. Note that Whole_AK has a higher starting recall but a higher false positive rate in comparison to Whole_AK_Negative_Resistant, due to being trained with fewer negative images. As stages are progressively removed from each classifier, Whole_AK_Negative_Resistant maintains a lower false positive rate (FPR), indicating that increased training with more negative images per stage can produce a more discriminative classifier. By increasing the total training images from 2000 to 5660

per stage, Whole_AK_Negative_Resistant recall was decreased by 6.0% and lowered the FPR from $2.39278 * 10^{-7}$ to $1.45846 * 10^{-7}$. In order to achieve an a recall rate of 85%, Whole_-AK required 6 classifier stages to be removed, while Whole_AK_Negative_Resistant required 7 classifier stages to be removed.
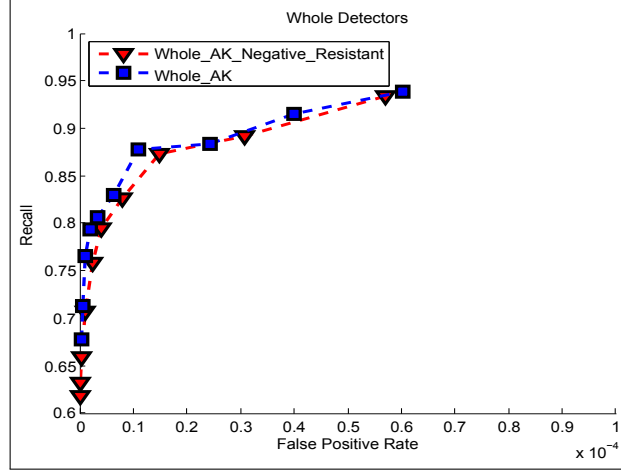
Figure 5.1: Whole Image Detections.
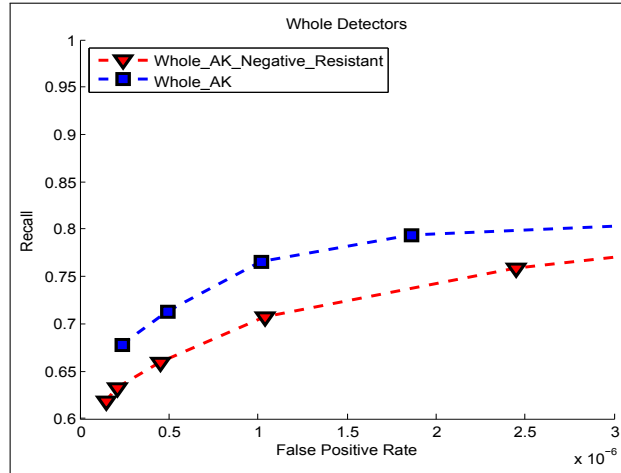
Figure 5.2: Whole Image Detections Capped at a False Positive Rate of $3 * 10^{-6}$.

5.2 AK-47 Detection with Part Trained Classifiers

Classifiers were trained for both left and right halves of an AK-47 to test the hypothesis that part trained classifiers have increased recall in relation to whole trained classifiers. Both the Left_Half classifier and the Right_Half classifier were trained with 5660 negative samples per stage in order to lower the false positive rate. It was also hypothesized that part classifiers would

increase the false positive rate. Training for each classifier required approximately two days on an Intel Core 2 CPU at 2.4 GHz with 2 GB of RAM, with the Windows XP operating system.

5.2.1 Part Trained Classifier Results

Results for part detectors are shown in Figure 5.3 comparing the Left_Half classifier against the Right_Half classifier. This is a comparison of detecting AK-47s through just detection of one of the parts. The Left_Half detector (all stages) has a recall of 73.6%, but generates 208.2 FPM. The Left_Half detector (all stages) increases recall over Whole_AK_Negative_Resistant by 11.8%, but also increases the false positive rate to $6.56062 * 10^{-7}$ (up from $1.45846 * 10^{-7}$ with the Whole_AK_Negative_Resistant classifier). The Right_Half detector (all stages) has a recall of 78.4%, but generates 1414.8 FPM. The Right_Half detector (all stages) increased recall over the Whole_AK_Negative_Resistant classifier by 16.6% but had a large increase in the false positive rate to $4.45696 * 10^{-6}$. In order to achieve an a recall rate of 85%, Left_Half required 6 classifier stages to be removed, while Right_Half required 3 classifier stages to be removed.

While part detectors have higher recall than whole detectors, part detectors alone generate far more false positives than whole trained techniques. The Right_Half detector has the highest starting recall of any classifier, indicating that AK-47s can be effectively detected by searching for the barrel, but results in too many false positives for incorporation into an operational system. The results also indicate that part trained classifiers have increased recall over whole trained classifiers, and show promise for increasing recall if the false positive rate can be controlled by another mechanism.
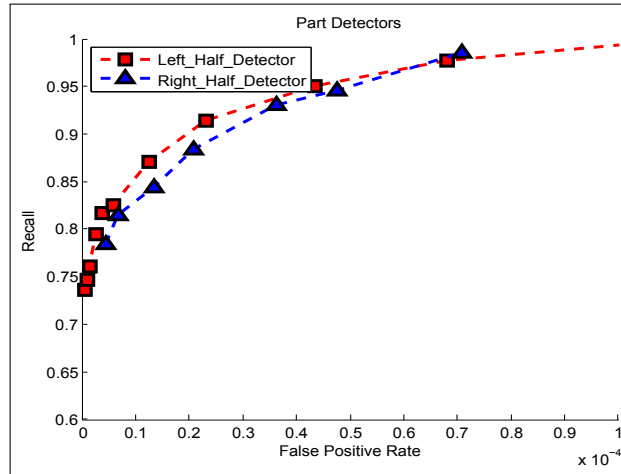

Figure 5.3: Part Image Detections.

31

5.3 Parts-based AK-47 Detection with an SVM for Structural Model Classification

Given the results from the part trained classifiers, it was hypothesized that the part classifiers could be used to increase recall, with a Support Vector Machine (SVM) used to control the false positive rate by ensuring that part detections matched a learned model. The structural model for AK-47 detection was introduced in Chapter 3. Training the SVM required approximately 2 seconds on Intel Core 2 CPU at 2.0 GHz with 4 GB of RAM, with the Windows Vista operating system.

5.3.1 Parts-based AK-47 Detection with an SVM for Structural Model Classification Results

The results for parts-based AK-47 with an SVM structural model classification are contained in Figure 5.4. The Left_Half and Right_Half classifiers were each individually applied to the image, with a Support Vector Machine used to evaluate all combinations of detections. In order to achieve a recall rate approaching 85%, stages must be removed from both classifiers.

The Left_Half detector (all stages) and Right_Half detector (all stages) using a Support Vector Machine increases the recall over the Whole_AK_Negative_Resistant classifier, while significantly decreasing the FPM over whole and part-based detectors. The Left and Right Half detectors with SVM had a starting recall of 69.1%, while producing only 17.5 FPM. In comparison against the Whole_AK_Negative_Resistant classifier, the parts-based SVM classifier increased recall by 7.3%, with a 57.5% reduction in the amount of false positives per minute. At the upper end of the recall scale, as both stages are progressively removed from both classifiers, the false positive rate greatly increases. This is due to the combination of all left and right detections being applied and checked against the SVM model. This indicates that if the part classifiers are not very discriminative, then the combination of all left and right detections in an image can increase the chance of classifying detections as a false positive. In order to achieve a recall rate of 85%, 7 classifier stages must be removed from both of the Left_Half and Right_Half classifiers.

5.4 Parts-based AK-47 Detection with an MLP for Structural Model Classification

A Multilayer Perceptron was also trained in order to compare against the Support Vector Machine. Once again, it was hypothesized that part classifiers could be used for increased recall,
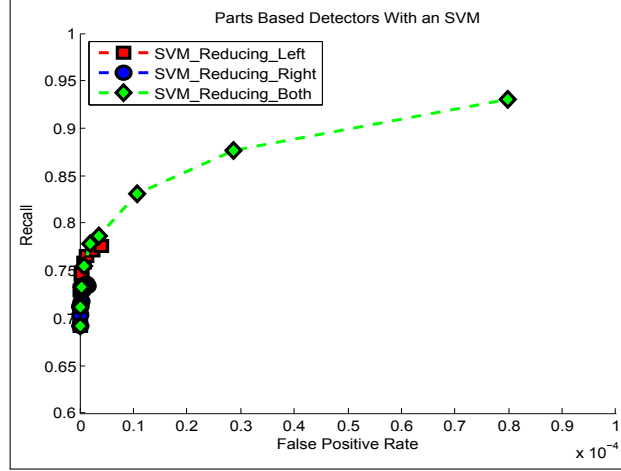
Figure 5.4: Parts-based SVM Image Detections.
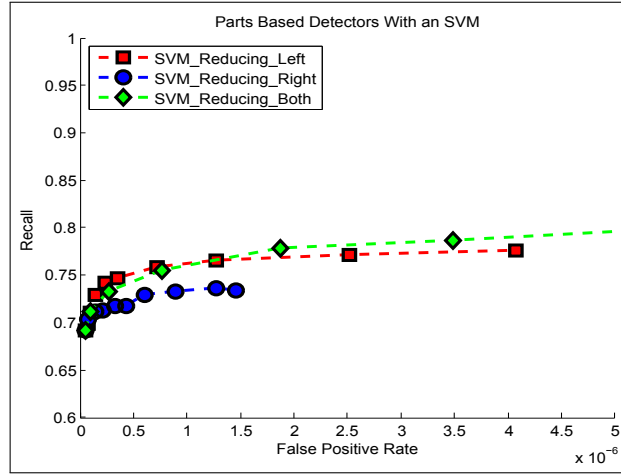
Figure 5.5: Parts-based SVM Image Detections Capped at $5 * 10^{-6}$.

while using the Multilayer Perceptron to control the false positive rate by comparing part detections against the AK-47 structural model developed in Chapter 3. Training the MLP required approximately 2 seconds on Intel Core 2 CPU at 2.0 GHz with 4 GB of RAM, with the Windows Vista operating system.

5.4.1 Parts-based AK-47 Detection with an MLP for Structural Model Classification Results

Results for the image set with part detectors and an MLP is contained in Figure 5.6. The Left -
Half and Right Half classifiers were applied to the image, with a Multilayer Perceptron used

33

to evaluate all combinations of detections. In order to achieve a recall rate approaching 85%, stages must be removed from both classifiers.

The Left_Half and Right_Half detectors (all stages) using a Multilayer Perceptron also increased the recall over the Whole_AK_Negative_Resistant classifier, while significantly decreasing the FPM over whole and parts-based detectors. The Left and Right Half detectors with MLP had a starting recall of 68.8%, while producing only 16.3 FPM. In comparison against Whole_AK_-Negative_Resistant classifier, the parts-based MLP classifier increased recall by 7.0%, with a 60.4% reduction in the amount of false positives per minute. Once again, at the upper end of the recall scale, as both stages are progressively removed from both classifiers, the false positive rate greatly increases. This is due to the combination of all left and right detections being applied and checked against the MLP model. This indicates that if the part classifiers are not very discriminative, then the combination of all left and right detections in an image can increase the chance of classifying detections as a false positive. In order to achieve a recall rate of 85%, 7 classifier stages must be removed from both of the Left_Half and Right_Half classifiers.
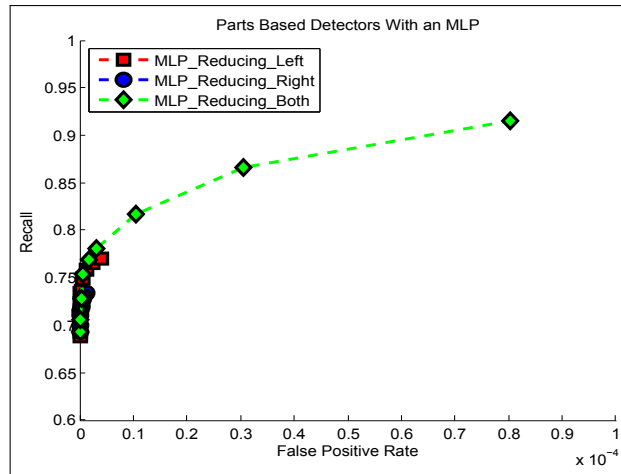
Figure 5.6: Parts-based MLP Image Detections.

5.5 All Detector Performance - Recall vs. FPR

In Figure 5.8, all detectors all compared on a single graph. Due to the wide variation in false positive rates generated, detector performance is best evaluated where a chosen operational system would likely operate. In Figure 5.9, the graph is capped at an FPR of $2 * 10^{-7}$ false positives per area checked, or about one false positive per 5 million areas checked.
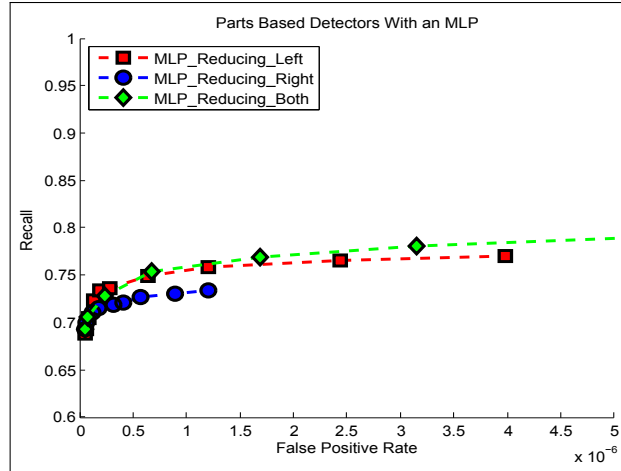
Figure 5.7: Parts-based MLP Image Detections Capped at $5 * 10^{-6}$.
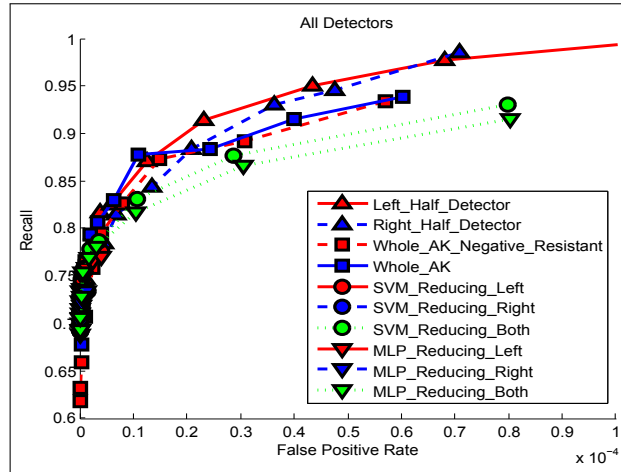
Figure 5.8: All Detector Performance.

5.6 All Detector Performance: Recall vs. False Positives per Minute of Video

In order to better represent the false positive rate of an operational system, the FPR is also shown in terms of the number of false positives per minute of video at a 15 frames a second at 640 x 480 resolution. See Figure 5.10. In Figure 5.11, the graph is capped at 200 false positives per minute of video to show classifier performance at the lowest false positive rates.
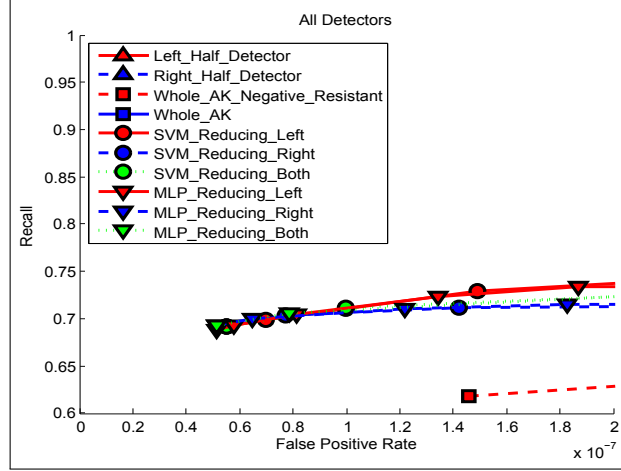
Figure 5.9: All Detector Performance Capped at FPR of $2 * 10^{-7}$.
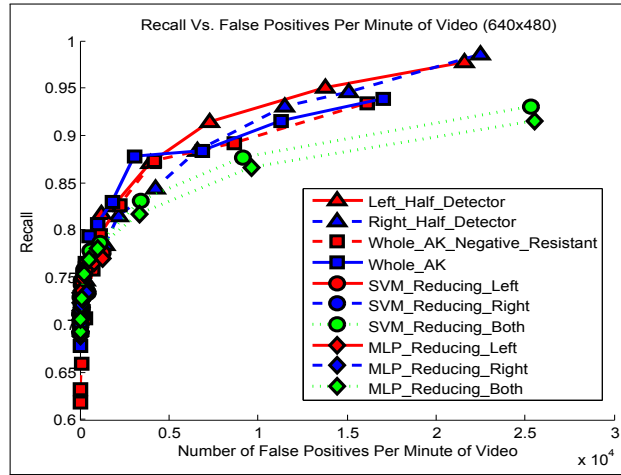
Figure 5.10: All Detectors. Recall vs. False Positives per Minute of Video.

5.7   Overall Analysis of Classifiers for AK-47 Detection

Whole object classifiers trained on AK-47s using traditional Viola-Jones techniques are capable of detecting weapons in images and video. While suitable, whole object classifiers do suffer from several problems when trying to detect weapons. First, weapon silhouettes become difficult to distinguish when the background is sufficiently cluttered, leading to the Viola-Jones cascade to reject the subwindow prior to reaching the end. This results in lower recall for the classifier. Next, since each subwindow is evaluated independently and the number of subwindows in an image can be quite large, the false positive rate is typically higher. This is due to the fact that there is no other mechanism with which to evaluate the subwindow to ensure that the
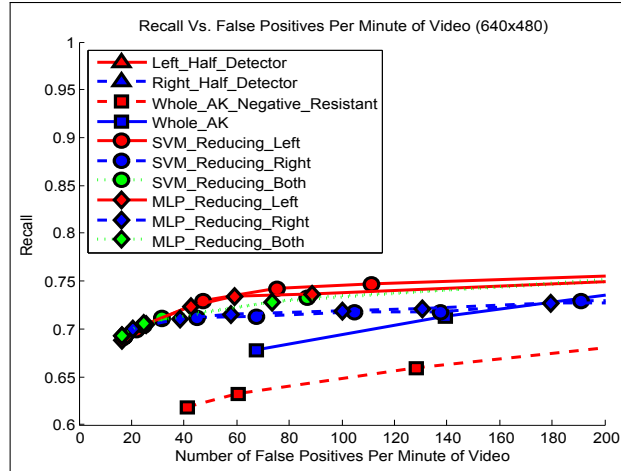
Figure 5.11: All Detector Performance, Recall vs. FPM, capped at 200 False Positives per Minute of Video.
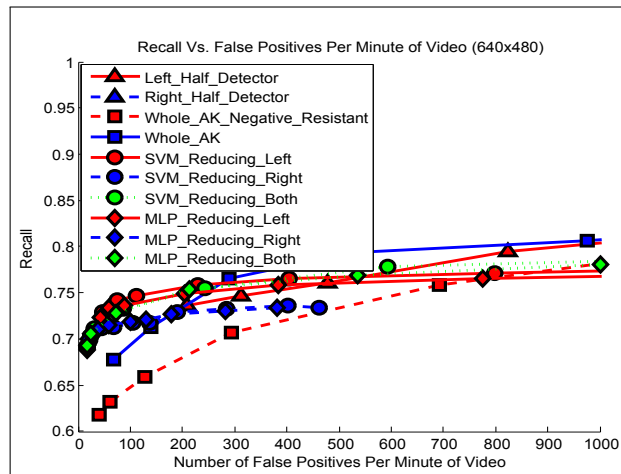
Figure 5.12: All Detector Performance, Recall vs. FPM, capped at 1000 False Positives per Minute of Video.

classifier detection is consistent with other object detections in the image.

Part classifiers outperform all other detectors in regards to recall, due to being trained on a simpler shape; however, the false positive rates generated by part detectors are too high to be incorporated into an intelligence application. First, the Right_Half detector demonstrates that a large number of AK-47s can be detected by simply searching for the barrel. While effective in regards to recall, the Right_Half detector is unsuitable for a surveillance system since there are far too many occurences of horizontal lines in an environment. For the Left_Half detector, the shape of the rear end of the AK-47 is slightly more complex. This means that the Left_Half detector is trained to recognize a shape with greater information, since the rear end of the AK-47

encompasses a pistol grip and a distinctive magazine shape, rather than horizontal lines for the barrel. This is evidenced by the lower recall and lower false positive rate of the the Left_Half detector in comparison with the Right_Half detector. Recall for the Left_Half detector is greater than whole detectors, since this is a simpler shape than the whole object.

Results from the part detectors support the hypothesis that recall is increased by searching for simpler shapes, while also leading to an increase in the number of false positives. Parts-based techniques offer potential for controlling the rising false positive rate, as long as the part detections are somewhat discriminative. The best overall performer in terms of the false positive rate for a simulated surveillance system (Figure 5.11) is the Left_Half and Right_Half detectors (all stages) using a Multilayer Perceptron to classify part detections in relation to the AK-47 Structural Model (Figure 3.8). Left_Half and Right_Half detectors (all stages) using a Support Vector Machine also produce favorable results. Both techniques lead to an increase in recall and a significant reduction in the amount of false positives detected in comparison with the best whole trained classifiers for use in a simulated surveillance system. Two important considerations are evident in the result graphs. First, due to the Structural Model, recall of an entire object is contingent on detection of both a left and right half of an AK-47. Reducing just one of the two classifiers can provide a slight increase in recall, but performance reaches a plateau very rapidly, since while one part is being detected, the other is not detected at all. This means that the individual part detection is then discarded, leading to a missed detection. Another consideration indicated by the resulting graphs is that parts-based techniques only perform better than whole classifiers, as long as the parts classifiers are somewhat discriminative. As classifier stages were removed from both left and right classifiers, the amount of false part detections in an image greatly increases. Since the parts-based technique classifies all combinations of left and right half detections against the Structural Model, this increases the likelihood of a false detection of an entire AK-47.

The techniques employed in this thesis are suitable for video and still images, though the classifier choice may differ depending on the application. For video surveillance, parts-based techniques offer substantial benefits for AK-47 detection in comparison against either whole or part classifiers. For video surveillance, a discriminitive classifier choice is needed due to the extremely large amount of subwindows checked in each minute of video. While temporal processing was not conducted to obtain any results in this thesis, the parts-based techniques can increase recall, while significantly reducing the amount of false positives, provided few, if any, classifier stages are removed from the left and right half detectors. By using temporal process-

ing, an AK-47 might not be detected in a particular frame, but may be detected in subsequent frames. Temporal processing over a video sequence is likely to further increase recall, while still limiting the false positive rate (see Future Work in Chapter 6). For still images, higher rates of recall may be required due to only having one chance to detect the weapon. Parts-based techniques continue to offer significant advantages for still frames over whole and part classifiers, if a recall of less than approximately 75% is acceptable. For desired rates of 75% and greater, whole trained classifiers begin to offer advantages over parts-based methods, due to the lack of discriminative part combinations.

The results of all classifiers in terms of reported amounts of false positives per minute of video contains an indication of the utility of each classifier. Instead of having to watch a minute of video, an analyst will be required to scan only cropped image areas for actual AK-47s. In comparison to whole object detectors, unreduced left and right half classifiers with a Multilayer Perceptron have a 60.4% reduction in the amount of false positives with a 7.0% increase in recall, enabling an analyst to rapidly scan images in less time required to watch an entire minute of video.

5.8 Detecting AK-47s at a Variety of In-plane Rotated Angles

The above tests were conducted on an image set with right facing AK-47s, with angles similar to those used in training. By utilizing part trained classifiers with either an SVM or MLP trained to recognize an AK-47 at a particular angle, AK-47s at other angles can be found by rotating the images. With an increase in the amount of runs for the detector, the false positive rate was expected to increase. In order to test this hypothesis, a prototype was tested over a separate image set of 3727 images from Dr. Garfinkel's govdocs1 corpus [47]. This prototype test incorporated Left and Right Half detectors (all stages) and an SVM, with a rotation angle of 10 degrees. None of the images contained AK-47s. Out of 1,517,722,847,280 areas checked, 571 false positives were detected, with a FPR of $3.76222 * 10^{-10}$ or approximately one false positive per 2,658,005,114 areas checked. While the detector has not been tested against a large positive image set at this point (due to a lack of a large image set with in-plane rotated AK-47s), Figure 5.13 confirms the capability to find rotated AK-47s with these methods of training. Classifiers trained with these methods also can be used to find weapons similar to AK-47s, including AK-74s and RPKs (Figure 5.14).

Figure 5.13: Parts Based Viola-Jones Classifiers With SVM True Positives. Images are Publicly Available at [48, 49].

Figure 5.14: Weapons Similar to an AK-47 (including AK-74s and RPKs) can be Found With Classifiers and Methods Used in This Thesis. Images are Publicly Available at [2, 50].

CHAPTER 6:
Conclusions

6.1 Concluding Remarks

Our experiments show that parts-based Viola-Jones classifiers combined with either a Support Vector Machine or Multilayer Perceptron leverage the high recall capability of part detectors and significantly reduce false positives in comparison to both the individual parts by themselves and whole object detectors, as long as the part classifier for the left and right halves are sufficiently discriminitive. Classifiers trained to detect parts of an AK-47 exhibit a high recall, but a poor false positive rate when compared against classifiers trained on the whole object. Our novel technique leverages the rapid detection inherent in Viola-Jones classifiers to detect AK-47s at a variety of scales, lighting, and background environments, while at the same time increasing recall and reducing false positives in comparison to traditional whole object approaches for video applications.

The techniques employed in this thesis are suitable for video and still images. For video surveillance, parts-based techniques offer substantial benefits for AK-47 detection in comparison against either whole or part classifiers, since a discriminitive classifier choice is needed due to the extremely large amount of subwindows checked in each minute of video. For still images, higher rates of recall may be required due to only having one chance to detect the weapon. Parts-based techniques continue to offer significant advantages for still frames over whole and part classifiers, if a recall of less than approximately 75% is acceptable. For desired rates of 75% and greater, whole trained classifiers begin to offer advantages over parts-based methods, due to the lack of discriminative part combinations.

The results of all classifiers in terms of reported amounts of false positives per minute of video contains an indication of the utility of each classifier. Instead of having to watch a minute of video, an analyst will be required to scan only cropped image areas for actual AK-47s.

This research directly benefits modern operational forces. Intelligence analysts are increasingly reliant on imaging systems, and require capabilities to deal with the growing amounts of data produced by surveillance, collection, and forensic systems. By rapidly locating an AK-47 in a video or image, analysts can focus on exploiting suspicious media and provide timely, relevant intelligence to forces in theater. Weapon detection in video also supports data fusion efforts and

collection management functions to better automate future persistent Intelligence, Surveillance, and Reconnaissance (ISR) systems. While initial results from these experiments demonstrate a capability, more work is needed to further increase recall and lower the amount of false positives detected.

6.2 Future Work

6.2.1 Training

All training images for this thesis were provided by videos found on the internet. Training by video provides the capability to rapidly assemble a database of images, though similar backgrounds in images likely impact the overall classifier training. A more diverse image set, with a variety of backgrounds will likely improve recall. During testing, images with false positives should be fed back to retrain the classifier in order to lower the false positive rate. Additionally, an operational system should incorporate actual images from captured hard drives and surveillance imagery in order to train for the likely domain.

6.2.2 Additional Part Training for AK-47 Detection

This thesis compared classifiers trained on a whole AK-47 against a parts-based technique using part combinations of left and right AK-47 halves. By training additional parts, the false positive rate may be lowered. Additional part detections may improve the likelihood of finding an object in situations where the sillouette of the object is obscured by back ground or when part of an object is occluded in the image.

6.2.3 Automated Part Training

Automated part training can provide several advantages over hard coding parts. First, humans do not know the best selection of parts to train. It is possible that by training parts through the use of interest point operators and clustering, a diverse set of parts can be trained and compared to find a better combination than those found through human selection. Second, automated part training is much faster to implement. As the number of parts increases, the work required to annotate parts for training and testing increases. Automated part training can decrease the workload required to train a variety of parts.

6.2.4 Probabalistic Cascade and Part Probability Distributions

While the binary Viola-Jones cascade is fast and efficient, each subwindow is evaluated independently of all other subwindows in the image. By using neighboring window detections

(including multi-scale detections), the overall likelihood of a detection can be increased. Part probability distributions can also be used to "vote" for object centers. With postprocessing turned off, the parts will likely be detected across multiple subwindows. With each part "voting" for a center, there will likely be a high probability generated at the object center. The binary nature of the Viola-Jones cascade also contributes to lower recall. An object is detected only after passing through all stages of the cascade. By returning a probability for each subwindow instead of a binary result, each subwindow can be evaluated and contributes to the overall probability of an object being present. The overall decision for object detection should be delayed until after all subwindows have been evaluated, with each window voting for a center with a likelihood of detection.

6.2.5   Training Multiple Object Classes Simultaneously

Due to the intelligence application of this research, classifiers must be robust and capable of detecting a variety of objects in cluttered, dynamic scenes. In [51], a technique for training multiple object classes simultaneously using boosting is discussed. Since most modern rifles share common characteristics, it would be far more efficient to train a class of objects, rather than a specific classifier for each weapon type.

6.2.6   Integration with Forensics/Surveillance Applications

The overall objective for this computer vision research is to enable detection of suspicious content in images, whether on a hard drive, website, or during surveillance operations from a variety of platforms. Web crawlers and content detection programs can utilize a library of trained classifiers to scan images, or subsample video frames for suspicious content. Trained classifiers can also be integrated with surveillance applications to reduce the burden of having a human constantly monitoring a video feed for suspicious activity. Temporal processing of detections can help to improve recall and lower the false positive rates since, an AK-47 might not be detected in a particular frame for a video, but might be detected in susequent frames.

6.2.7   Integration With Natural Language Processing

Computer vision techniques combined with Natural Language techniques may yield results when searching for suspicious media. Digital content containing both suspicious words and images increases the likelihood of a successful detection. Additionally, after object detection techniques identify suspicious media pages, words in proximity to the suspicious image can be used to refine language models, and possibily identify new vocabulary words that help in discriminating enemy activity.

THIS PAGE INTENTIONALLY LEFT BLANK

REFERENCES

[1] Too much information: Taming the uav data explosion, 2010. http://www.defenseindustrydaily. com/uav-data-volume-solutions-06348/. [Accessed: May 22, 2010].

[2] Osama bin laden image— Wikipedia, the free encyclopedia, 2010. http://en.wikipedia.org/wiki/ Osama_bin_Laden. [Accessed: August 23, 2010].

[3] P. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR01*, pp. 511–518, 2001.

[4] P. Viola and M. Jones. Robust real-time object detection. In *Second international workshop on statistical and computational theories of vision - modeling, learning, computing, and sampling*, 2001.

[5] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, pp. 900–903, 2002.

[6] S. Du, N. Zheng, Q. You, Y. Wu, M. Yuan, and J. Wu. Rotated haar-like features for face detection with in-plane rotation. *Interactive Technologies and Sociotechnical Systems*, pp. 128–137.

[7] Y. Deng and G. Su. Face Detection Based on Fuzzy Cascade Classifier with Scale-invariant Features. *International Journal of Information Technology*, 12(5), 2006.

[8] H. Takatsuka, M. Tanaka, and M. Okutomi. Distribution-Based Face Detection using Calibrated Boosted Cascade Classifier. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pp. 351–356, 2007.

[9] Z. Lin, G. Hua, and L.S. Davis. Multiple instance Feature for robust part-based object detection. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[10] N. Dalai, B. Triggs, I. Rhone-Alps, and F. Montbonnot. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 1, 2005.

[11] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. *Computer Vision–ECCV 2006*, pp. 428–441.

[12] Z. Qiang, Y. Mei-Chen, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *IEEE Computer Vision and Pattern Recognition Conference*, volume 2, pp. 1491–1498, 2006.

[13] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. PAMI, 2009, press.

[14] S. Kumar and M. Hebert. Multiclass discriminative fields for parts-based object detection. In *Snowbird Learning Workshop*, 2004.

[15] J. Zhang, Y. Yan, and M. Lades. Face recognition: eigenface, elastic matching, and neural nets. *Proceedings of the IEEE*, 85(9):1423, 1997.

[16] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1): 71–86, 1991.

[17] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Computer VisionECCV'96*, pp. 43–58.

[18] J. Wu, Z. Zhou, L. Zhou, and D. Hu. Learning spatial prior with automatically labeled landmarks. In *Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on*, volume 1, 2008.

[19] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (29):2007.

[20] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, pp. 994–1000, 2005.

[21] J.G. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7):1169, 1988.

[22] K. Mikolajczyk, A. Zisserman, and C. Schmid. Shape recognition with edge-based features. In *Proceedings of the British Machine Vision Conference, Norwich, UK*. Citeseer, 2003.

[23] J. Canny. A computational approach to edge detection. *Readings in computer vision: issues, problems, principles, and paradigms*, 184, 1987.

[24] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. *Computer Vision-ECCV 2004*, pp. 69–82, 2004.

[25] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, 11 (3):283–318, 1993.

[26] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, p. 50. Manchester, UK, 1988.

[27] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. 26(11):1475–1490, November 2004.

[28] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 878–885. IEEE Computer Society, 2005.

[29] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, volume 2, 2003.

[30] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, 2006.

[31] M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. In *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings*, volume 2, 2000.

[32] R. Kindermann, J.L. Snell, and American Mathematical Society. *Markov random fields and their applications*. American Mathematical Society Providence, Rhode Island, 1980.

[33] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (20), 1998.

[34] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa. Automated detection of human for visual surveillance system. In *Proc. of Intl. Conf. on Pattern Recognition*, pp. 865–869, 1996.

[35] J. Nam, M. Alghoniemy, and A. H. Tewfik. Audio-visual content-based violent scene characterization. In *International Conference on Image Processing (ICIP 98*, pp. 353–357, 1998.

[36] Nuno Vasconcelos and Andrew Lippman. Towards semantically meaningful feature spaces for the characterization of video content. In *Proceedings of IEEE Intl Conference on Image Processing*, pp. 542–545, 1997.

[37] Ankur Datta, Mubarak Shah, Niels Da, and Niels Da Vitoria Lobo. Person-on-person violence detection in video data. In *Proc. Intl Conference on Pattern Recognition*, pp. 433–438, 2002.

[38] D. Arsic, B. Schuller, and G. Rigoll. Suspicious behavior detection in public transport by fusion of low-level video descriptors. In *2007 IEEE International Conference on Multimedia and Expo*, pp. 2018–2021, 2007.

[39] T. Xiang and S. Gong. Video behavior profiling for anomaly detection. *IEEE transactions on pattern analysis and machine intelligence*, 30(5):893, 2008.

[40] How face detection works, 2010. http://www.cognotics.com/opencv/servo_2007_series/part_2/sidebar.html. [Accessed: April 13, 2010].

[41] Viola-jones object detection framework image— Wikipedia, the free encyclopedia, 2010. http://en.wikipedia.org/wiki/Viola-Jones_object_detection_framework. [Accessed: August 23, 2010].

[42] E. Alpaydin. *Introduction to machine learning*. The MIT Press, 2004.

[43] Support vector machine image— Wikipedia, the free encyclopedia, 2010. http://en.wikipedia.org/wiki/Support_vector_machine. [Accessed: April 13, 2010].

[44] Neural networks, 2010. http://opencv.willowgarage.com/documentation/cpp/neural_networks.html. [Accessed: April 11, 2010].

[45] R. Rojas. *Neural networks: a systematic introduction*. Springer, 1996.

[46] Ak-47 image— Wikipedia, the free encyclopedia, 2010. http://en.wikipedia.org/wiki/AK-47. [Accessed: August 23, 2010].

[47] S. Garfinkel, P. Farrell, V. Roussev, and G. Dinolt. Bringing science to digital forensics with standardized forensic corpora. *digital investigation*, 6:S2–S11, 2009.

[48] Ramadi militants image— Wikipedia commons, 2010. http://commons.wikimedia.org/wiki/File:Ramadi01.jpg.jpg. [Accessed: August 23, 2010].

[49] Pkk militant image— Wikipedia commons, 2010. http://commons.wikimedia.org/wiki/File:PKK_Militant.jpg. [Accessed: August 23, 2010].

[50] Rpk image— Wikipedia commons, 2010. http://commons.wikimedia.org/wiki/File:Soviet_RPK-74.JPEG. [Accessed: August 23, 2010].

[51] Antonio Torralba Kevin, Kevin P. Murphy, and William T. Freeman. Sharing features: Efficient boosting procedures for multiclass object detection. In *CVPR*, pp. 762–769, 2004.

APPENDIX A:
Tables

The section contains the results for all detectors. Each classifier is reported with its associated recall, false positives per minute of 640x480 video at 15 frames per second, and false positive rate. Please see Chapter 3 for the AK-47 structural model, and how performance measures were calculated.

A.1 Whole Object Detectors

Table A.1: Summary of Findings—Whole_AK

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Whole_AK_19	0.6783	67.68	2.39E-07
Whole_AK_18	0.7132	139.56	4.93E-07
Whole_AK_17	0.7656	289.12	1.02E-06
Whole_AK_16	0.7933	526.68	1.86E-06
Whole_AK_15	0.8064	975.36	3.45E-06
Whole_AK_14	0.8296	1817.60	6.43E-06
Whole_AK_13	0.8777	3107.88	1.10E-05
Whole_AK_12	0.8835	6864.60	2.43E-05
Whole_AK_11	0.9155	11303.05	4.00E-05
Whole_AK_10	0.9388	17062.07	6.03E-05
Whole_AK_9	0.9446	26173.94	9.25E-05
Whole_AK_8	0.9723	35109.83	0.12E-03

Table A.2: Summary of Findings—Whole_AK_Negative_-Resistant

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Whole_AK_Neg_Resist_18	0.6186	41.25	1.46E-07
Whole_AK_Negative_Resistant - Continued on next page			

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Whole_AK_Neg_Resist_17	0.6317	60.59	2.14E-07
Whole_AK_Neg_Resist_16	0.6593	128.28	4.53E-07
Whole_AK_Neg_Resist_15	0.7074	294.92	1.04E-06
Whole_AK_Neg_Resist_14	0.7583	692.35	2.44E-06
Whole_AK_Neg_Resist_13	0.7947	1172.94	4.15E-06
Whole_AK_Neg_Resist_12	0.8267	2244.36	7.93E-06
Whole_AK_Neg_Resist_10	0.8733	4202.50	1.49E-05
Whole_AK_Neg_Resist_9	0.8922	8700.58	3.08E-05
Whole_AK_Neg_Resist_8	0.9344	16161.49	5.71E-05
Whole_AK_Neg_Resist_7	0.9548	25637.27	9.06E-05
Whole_AK_Neg_Resist_6	0.9839	39249.15	0.14E-03

A.2   Part Detectors

Table A.3: Summary of Findings—Left_Half_Detector

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Left_Half_Unreduced	0.7365	208.26	6.56E-07
Left18	0.7467	311.78	9.82E-07
Left17	0.7612	477.04	1.50E-06
Left16	0.7947	823.53	2.59E-06
Left15	0.8165	1219.48	3.84E-06
Left14	0.8253	1882.67	5.93E-06
Left13	0.8704	3992.34	1.26E-05
Left12	0.9141	7323.96	2.31E-05
Left11	0.9505	13778.00	4.34E-05
Left10	0.9767	21597.13	6.80E-05
Left9	0.9941	32397.69	0.10E-03

Table A.4: Summary of Findings—Right_Half_Detector

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Right_Half_Unreduced	0.7845	1414.84	4.46E-06
Right18	0.8151	2151.75	6.78E-06
Right17	0.8442	4251.29	1.34E-05
Right16	0.8835	6627.90	2.09E-05
Right15	0.9301	11521.50	3.63E-05
Right14	0.9461	15083.19	4.75E-05
Right13	0.9854	22516.19	7.09E-05

A.3 Parts Based Classifiers with a Support Vector Machine

Table A.5: Summary of Findings—Parts Based Classifiers with an SVM

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Left(Unred)Right(Unred)	0.6914	17.50	5.52E-08
Right(Unreduced)Left18	0.6986	22.11	6.97E-08
Right(Unreduced)Left17	0.7103	31.63	9.97E-08
Right(Unreduced)Left16	0.7292	47.30	1.49E-07
Right(Unreduced)Left15	0.7423	75.25	2.37E-07
Right(Unreduced)Left14	0.7467	111.19	3.50E-07
Right(Unreduced)Left13	0.7583	229.45	7.23E-07
Right(Unreduced)Left12	0.7656	403.93	1.27E-06
Right(Unreduced)Left11	0.7714	798.34	2.51E-06
Right(Unreduced)Left10	0.7758	1294.12	4.08E-06
Right(Unreduced)Left9	0.7772	2011.37	6.34E-06
Right(Unreduced)Left8	0.7802	3450.18	1.09E-05
Parts Based Classifiers with an SVM - Continued on next page			

Table A.5—Summary of Findings—Parts Based Classifiers with an SVM—Continued

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Left18Right18	0.7117	31.63	9.97E-08
Left17Right17	0.7321	86.93	2.74E-07
Left16 Right16	0.7554	242.36	7.63E-07
Left15Right15	0.7787	593.76	1.87E-06
Left14Right14	0.7860	1107.67	3.49E-06
Left13Right13	0.8311	3389.05	1.06E-05
Left12Right12	0.8762	9140.59	2.88E-05
Left11Right11	0.9301	25360.63	7.99E-05
Left(Unreduced)Right18	0.7030	24.57	7.74E-08
Left(Unreduced)Right17	0.7117	45.15	1.42E-07
Left(Unreduced)Right16	0.7132	67.57	2.13E-07
Left(Unreduced)Right15	0.7176	105.05	3.31E-07
Left(Unreduced)Right14	0.7176	137.61	4.34E-07
Left(Unreduced)Right13	0.7292	191.06	6.02E-07
Left(Unreduced)Right12	0.7321	284.75	8.97E-07
Left(Unreduced)Right11	0.7365	403.01	1.27E-06
Left(Unreduced)Right10	0.7336	462.91	1.46E-06
Left(Unreduced)Right9	0.7278	555.67	1.75E-06

A.4   Parts Based Classifiers with a Multilayer Perceptron

Table A.6: Summary of Findings—Parts Based Classifiers
with an MLP

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Left(Unred)Right(Unred)	0.6885	16.28	5.13E-08
Right(Unreduced)Left18	0.6928	18.43	5.81E-08
Right(Unreduced)Left17	0.7045	25.80	8.13E-08
Right(Unreduced)Left16	0.7234	42.69	1.35E-07
Parts Based Classifiers with an MLP - Continued on next page			

Table A.6—Summary of Findings—Parts Based Classifiers with an MLP—Continued

Classifier Name	Recall	False Detections Per Minute	False Positive Rate
Right(Unreduced)Left15	0.7336	59.28	1.87E-07
Right(Unreduced)Left14	0.7365	88.77	2.80E-07
Right(Unreduced)Left13	0.7496	203.96	6.43E-07
Right(Unreduced)Left12	0.7583	383.04	1.21E-06
Right(Unreduced)Left11	0.7656	775.00	2.44E-06
Right(Unreduced)Left10	0.7700	1265.25	3.99E-06
Left18Right18	0.7059	24.88	7.84E-08
Left17Right17	0.7278	73.72	2.32E-07
Left16Right16	0.7540	213.79	6.73E-07
Left15Right15	0.7685	535.40	1.69E-06
Left14Right14	0.7802	1002.00	3.16E-06
Left13Right13	0.8165	3355.88	1.10E-05
Left12Right12	0.8660	9661.86	3.04E-05
Left11Right11	0.9155	25511.45	8.04E-05
Left(Unreduced)Right18	0.7001	20.58	6.48E-08
Left(Unreduced)Right17	0.7103	38.70	1.22E-07
Left(Unreduced)Right16	0.7147	58.05	1.83E-07
Left(Unreduced)Right15	0.7190	100.13	3.15E-07
Left(Unreduced)Right14	0.7205	130.54	4.11E-07
Left(Unreduced)Right13	0.7263	179.38	5.65E-07
Left(Unreduced)Right12	0.7307	283.21	8.92E-07
Left(Unreduced)Right11	0.7336	381.20	1.20E-06

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Marine Corps Representative
   Naval Postgraduate School
   Monterey, California

4. Director, Training and Education, MCCDC, Code C46
   Quantico, Virginia

5. Director, Marine Corps Research Center, MCCDC, Code C40RC
   Quantico, Virginia

6. Marine Corps Tactical Systems Support Activity (Attn: Operations Officer)
   Camp Pendleton, California